

# Regularized Spectral Matching for Blind Source Separation. Application to fMRI Imaging.

Hichem Snoussi<sup>a\*</sup>, Vince D. Calhoun<sup>b</sup>

<sup>a</sup> IRCCyN, Institut de Recherche en Communications et Cybernétiques de Nantes, Ecole Centrale de Nantes, 44321, Nantes, France, email: snoussi@irccyn.ec-nantes.fr Phone: 0033 2 40376925, Fax: 0033 2 40376930.

<sup>b</sup> Olin Neuropsychiatry Research Center, Institute of Living, Hartford, CT 06106, email: vince.calhoun@yale.edu Phone: 1-860-545-7768, Fax: 1-860-545-7797

*Abstract* —

The main contribution of this paper is to present a Bayesian approach for solving the noisy instantaneous blind source separation problem based on second order statistics of the time varying spectrum. The success of the blind estimation relies on the non stationarity of the second order statistics and their inter-source diversity. Choosing the time frequency domain as the signal representation space and transforming the data by a short time Fourier transform (STFT), our method presents a simple EM algorithm which can efficiently deal with the time varying spectrum diversity of the sources. The estimation variance of the STFT is reduced by averaging across time frequency sub-domains. The algorithm is demonstrated on a standard functional resonance imaging (fMRI) experiment involving visual stimuli in a block design. Explicitly taking into account the noise in the model, the proposed algorithm has the advantage of extracting only relevant task related components and considers the remaining components (artifacts) as noise.

*Keywords* — Blind Source Separation, EM algorithm, Short Time Fourier Transform, Maximum Likelihood, fMRI imaging.

## I. INTRODUCTION

Since the beginning of the last decade, extensive research has been devoted to the problem of blind source separation (BSS). The attractiveness of this particular problem is essentially due to both its applicative and theoretical challenging aspects. This research has given rise to the development of many methods aiming to solve this problem (see [1, 2] for an overview). An interesting aspect of this emerging field, still open to more research, is the fact that the theoretical development evolves in pair with the real world application specificities and requirements. Extracting components and time courses of interest from fMRI data [3, 4] is a representative illustration of this statement. BSS can be analyzed with two dual approaches: source separation as a source reconstruction problem or source separation as a decomposition problem. In the first approach, one assumes that, during an experiment  $\mathcal{E}$ , the collected data  $\mathbf{x}_{1..T} = \{\mathbf{x}_1, \dots, \mathbf{x}_T\}$  are not a faithful copy of the original

process of interest  $\mathbf{s}_{1..T}$  under study (sources). In other words, the observed data  $\mathbf{x}_{1..T}$  are some transformation  $\mathcal{F}$  of the sources  $\mathbf{s}_{1..T}$  corrupted with a stochastic noise  $\mathbf{n}_{1..T}$  reflecting either the modeling incertitude or the superposition of real undesirable signals:

$$\mathbf{x}_{1..T} = \mathcal{F}(\mathbf{s}_{1..T}) \odot \mathbf{n}_{1..T},$$

where  $\odot$  is the operator modeling the noise superposition. Given the data  $\mathbf{x}_{1..T}$ , our objective is the recovery of the original sources  $\mathbf{s}_{1..T}$ . The second approach for dealing with the source separation problem is to consider it as a decomposition on a basis enjoying some particular statistical properties. For instance, PCA (Principal Component Analysis) relies on the decorrelation between the decomposed components and ICA (Independent Component Analysis) relies on their statistical independence. The decomposition approach can be considered as dual to the reconstruction approach (see Figure 1) as the existence of an original process is not required.

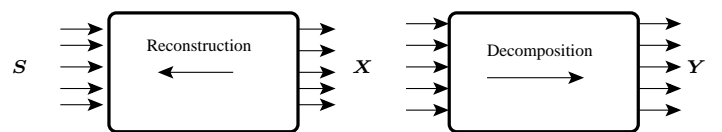


Fig. 1. Duality of reconstruction and decomposition approaches

In this paper, we consider the reconstruction approach with the noisy linear instantaneous mixture:

$$\mathbf{x}_t = \mathbf{A}\mathbf{s}_t + \mathbf{n}_t, \quad t = 1, \dots, T,$$

The challenging aspect of the BSS problem is the absence of any exact information about the mixing matrix  $\mathbf{A}$ .

Based on i.i.d source modeling, many proposed algorithms are designed to linearly demix the observations  $\mathbf{x}_{1..T}$ . The separation principle in these methods is based on the statistical independence of the reconstructed sources (Independent Component Analysis) [5–9]. However, ICA

is designed to efficiently work in the noiseless case. In addition, with the i.i.d assumption, the separation necessarily relies on high order statistics and treating the noisy case with the maximum likelihood approach leads to complicated algorithms [10–12].

Discarding the i.i.d assumption, source separation can be achieved with second order statistics. For instance, second order correlation diversity in the time domain [13], frequency domain [14] or time frequency domain [15] are successfully used to blindly separate the sources. Non stationary second order based methods are also proposed in [16–20] (see [21] and the references therein for a synthetic introduction of these concepts). Stationarity and non stationarity can approximately be seen as dual under Fourier transformation. For instance, based on the circular approximation, it is shown [22] that a finite sample correlated temporal stationary signal has a Fourier transform with non stationary decorrelated samples. We have recently proposed a maximum likelihood method to separate noisy mixture of Gaussian stationary sources exploiting this temporal / spectral duality [23, 24]. The Gaussian model of sources allows an efficient implementation of the EM (Expectation-Maximization) algorithm [25]. In this contribution, we extend this approach to deal with non stationary sources and a limited sample size of collected observations. Relying on the maximum likelihood principle and the Short Time Fourier Transform (STFT), our approach can be interpreted as a matching between a smoothed estimate of the spectrum of the observations and the theoretic structured spectrum arising from the mixture structure of data. Thus, we improve the previously proposed method [23, 24] not only by extending to time varying spectrum matching but also by regularizing (smoothing) the data spectrum. This regularization is of capital importance when the number of data is limited as in some fMRI experiments where the sample size corresponds to the duration of the scanning.

Our statistical spectral approach is different from the work proposed in [15] which is based on the joint diagonalization of several observations ambiguity matrices. In fact, our method is based on the maximum likelihood principle which is flexible to incorporate, using the Bayesian rule, any prior information about the mixing operator and the source spectra. Taking into account, in a consistent way, the prior information can be seen as a regularization of the source separation problem in difficult situations as in the underdeterminate, noisy, convolutive or time varying mixtures. Our method is based on the estimation of the mixing matrix, the source spectra and the noise covariance matrix. Thus, the same algorithm is applied to overdeterminate and underdeterminate cases without a prewhitening step. The method implicitly incorporates a denoising procedure and it is consequently robust to high level noise. The equations involved in the EM algorithm are very simple to implement. An interesting property of the proposed solution is the exploitation of second order spectral non stationarity. In addition, by partitioning the time frequency plane into horizontal bands, the marginal spectrum (integrating over

time the spectrograms) corresponds to an improved version of the separation of noisy stationary sources [23]. In fact, the smoothed periodograms, obtained by marginalization, are used instead of the empirical periodograms (corresponding to the Wigner-Ville distribution marginals).

The paper is organized as follows. In Section II, we briefly recall the maximum likelihood separation of noisy stationary sources. Section III is devoted to the main contribution of this paper. We develop the EM algorithm implementing the maximum likelihood solution in the time frequency domain. The ML criteria is interpreted as a Kullback-Leibler matching between smoothed Wigner-Ville spectra. We show how spectral non stationarity exploitation can be obtained by marginalizing the STFT representation. In Section IV, application to real fMRI signals illustrates the effectiveness of our proposed method comparing to the ICA solution.

## II. STATIONARY CASE

In this section, we recall the basics of the separation of noisy stationary sources and why the spectral domain is very appropriate in this case. An improvement of this method, based on marginalizing the time frequency spectrum, even in the stationary case, will be presented in section III.

We consider the noisy linear instantaneous mixture model<sup>1</sup>:

$$\mathbf{x}_t = \mathbf{A}\mathbf{s}_t + \mathbf{n}_t, \quad (1)$$

where  $\mathbf{x}_t$  is the  $(m \times 1)$  vector of observations at the time  $t$ ,  $\mathbf{s}_t$  is the  $(n \times 1)$  vector of sources,  $\mathbf{A}$  is the  $(m \times n)$  mixing matrix and  $\mathbf{n}_t$  the noise vector assumed to be white and stationary (with unknown covariance  $\mathbf{R}_e$ ).

A useful approximation is to consider the covariance matrix  $\mathbf{\Sigma}$  of a stationary signal  $x_{1..N} = [x_1, \dots, x_N]$  (for each channel) as circular. This approximation consists in circularizing the raw vector  $x_{1..N}$  (see Figure 2). Then, the matrix  $\mathbf{\Sigma}$  can be diagonalized in the Fourier basis with eigenvalues coinciding with the spectrum of the signal  $x_{1..N}$ :

$$\mathbf{\Sigma} \approx \mathbf{F}^* \mathbf{\Delta} \mathbf{F}, \quad \mathbf{\Delta} = \begin{pmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_N^2 \end{pmatrix} \quad (2)$$

where "\*" denotes the conjugate transpose and "T" denotes the transpose.  $\mathbf{F}$  is the Fourier matrix:

$$\{F_{ik} = \exp[-2j\pi(i-1)(k-1)/N]\}_{i=1..N}^{k=1..N}$$

As a consequence of equation (2), Fourier transforming the signal ( $\tilde{x}_{1..N} = x_{1..N} \mathbf{F}^T$ ) leads to a decorrelated signal with a diagonal covariance equal to the spectrum of the signal. Considering only the second order statistics, this is equivalent to a non stationary white Gaussian process. If the signal  $x_{1..N}$  is initially stationary white, then the

<sup>1</sup>Unless explicitly specified, the mixture is considered noisy linear instantaneous with no restriction to either overdeterminate or underdeterminate cases.

spectrum is constant. However, if the signal  $x_{1..N}$  has a correlation structure, the spectrum is non constant. This idea is exploited in the next paragraph II-A, to blindly estimate the mixing matrix, the source spectra and the noise covariance.

### A. Maximum Likelihood Criteria

Transforming the data by Fourier decomposition, the mixture model, at each Fourier mode  $k$  of the spectral domain, is,

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k) + \mathbf{n}(k)$$

where, for the sake of clarity, we use the same notation for the Fourier transformed variables. In the following, by  $\mathbf{x}_{1..N}$  and  $\mathbf{s}_{1..N}$  we refer to the whole sample of transformed observations and sources.

The sources are modeled by a non stationary white Gaussian process (real and imaginary components are independent). At each frequency  $k$ , the sources  $\mathbf{s}_k$  has a zero mean Gaussian distribution with diagonal covariance  $\mathbf{P}_k = \mathbb{E}[\mathbf{s}_k \mathbf{s}_k^*]$  (the diagonality reflects the source independence),

$$\mathbf{s}_k \sim \mathcal{N}(0, \mathbf{P}_k).$$

The diagonal elements  $[P_{cc}(k), k = 1..N]$  of the matrices  $\mathbf{P}_k$  are the power spectra of the sources. The noise is assumed to be zero mean white Gaussian with constant spectrum  $\mathbf{R}_\epsilon = \mathbb{E}[\mathbf{n}_k \mathbf{n}_k^*]$ . We also assume that the matrix  $\mathbf{R}_\epsilon$  is diagonal and that the diagonal elements may have different values (different noise detector levels).

We consider the maximum likelihood approach to jointly estimate the mixing matrix  $\mathbf{A}$ , the source spectral densities  $\mathbf{P}_k$  and the noise covariance  $\mathbf{R}_\epsilon$ , based on the mixture model (1) and the stationarity property:

$$(\hat{\mathbf{A}}, \hat{\mathbf{R}}_\epsilon, \{\hat{\mathbf{P}}_k\}) = \arg \max_{\mathbf{A}, \mathbf{R}_\epsilon, \{\mathbf{P}_k\}} p(\mathbf{x}_{1..K} | \mathbf{A}, \mathbf{R}_\epsilon, \{\mathbf{P}_k\}) p(\mathbf{A}, \mathbf{R}_\epsilon, \{\mathbf{P}_k\})$$

where  $p(\mathbf{A}, \mathbf{R}_\epsilon, \{\mathbf{P}_k\})$  contains the *a priori* information about the parameters to estimate.

The gaussianity (in other words, the restriction to second order statistics) and the spectral independence of sources and noise lead to explicit expression of the likelihood of  $\boldsymbol{\theta} = (\mathbf{A}, \mathbf{R}_\epsilon, \{\mathbf{P}_k\})$ :

$$\begin{aligned} p(\mathbf{x}_{1..N} | \boldsymbol{\theta}) &= \int p(\mathbf{x}_{1..N} | \mathbf{s}_{1..N}, \mathbf{A}, \mathbf{R}_\epsilon) p(\mathbf{s}_{1..N} | \mathbf{P}_{1..N}) d\mathbf{s}_{1..N} \\ &= \prod_k \int p(\mathbf{x}_k | \mathbf{s}_k, \mathbf{A}, \mathbf{R}_\epsilon) p(\mathbf{s}_k | \mathbf{P}_k) d\mathbf{s}_k \\ &= \prod_k |2\pi \mathbf{R}_k|^{-1} \exp\left[-\text{Tr}\left(\mathbf{R}_k^{-1} \mathbf{x}_k \mathbf{x}_k^*\right)\right]. \end{aligned} \quad (3)$$

where  $\mathbf{R}_k = \mathbf{A}\mathbf{P}_k\mathbf{A}^* + \mathbf{R}_\epsilon$  is the spectral covariance matrix of the observations  $\mathbf{x}_k$  at the mode  $k$ .

Thus, the method is implemented in the spectral domain exploiting the decorrelation property (circular approximation) and the spectral diversity of the sources [23], [24], [26]. The optimization is performed by an EM algorithm as the parameter identification can be described as an incomplete data problem, where  $\mathbf{x}_{1..K}$  are considered as incomplete data and  $\mathbf{s}_{1..N}$  are considered as hidden variables. As it will be shown later, the EM algorithm yields, implicitly, the Wiener filtering of the sources. Therefore, after convergence, we have an estimation of the parameter  $\boldsymbol{\theta} = (\mathbf{A}, \mathbf{R}_\epsilon, \{\mathbf{P}_k\})$  and a reconstruction of the sources  $\mathbf{s}_{1..N}$ .

### B. Interpretation of the ML criteria

We note that the number of the power spectrum coefficients  $\mathbf{P}_{1..N}$  to be estimated is  $N$ , the size of the whole sample. For this reason, we partition the spectral interval  $[1..N]$  into  $L$  sub-intervals  $\{\mathcal{D}_l\}_{l=1}^L$  (as in the noiseless case [20]). The choice of this partition depends on some *a priori* knowledge about the source components (see Figure 3). Then, we reduce the number of spectral coefficients to be estimated to  $L$ . Consequently, the observation covariance is constant in each sub-interval:  $\mathbf{R}_k = \mathbf{R}_l = \mathbf{A}\mathbf{P}_l\mathbf{A}^* + \mathbf{R}_\epsilon, \forall k \in \mathcal{D}_l$ .

The likelihood expression (3) can be rewritten by re-partitioning the modes,

$$p(\mathbf{x}_{1..K} | \boldsymbol{\theta}) = \prod_{l=1}^L |2\pi \mathbf{R}_l|^{-w_l} \exp\left[-\text{Tr}\left(\mathbf{R}_l^{-1} \sum_{k \in \mathcal{D}_l} \mathbf{x}_k \mathbf{x}_k^*\right)\right] \quad (4)$$

where  $w_l = |\mathcal{D}_l|$  is the number of modes belonging to the sub-interval  $\mathcal{D}_l$ .

Introducing the empirical observation covariances  $\hat{\mathbf{R}}_l = \sum_{k \in \mathcal{D}_l} \mathbf{x}_k \mathbf{x}_k^* / w_l$ , the log-likelihood can be considered as a weighted sum of Kullback-Leibler divergences between the structured spectral theoretical covariance and the spectral empirical covariance of the sub-intervals  $\mathcal{D}_l$  [20]:

$$\begin{aligned} \log p(\mathbf{x}_{1..K} | \boldsymbol{\theta}) &= -\sum_{l=1}^L w_l \left(-\ln |\mathbf{R}_l^{-1} \hat{\mathbf{R}}_l| + \text{Tr}\left(\mathbf{R}_l^{-1} \hat{\mathbf{R}}_l\right) - N_d\right) + \text{const} \\ &= -\sum_{l=1}^L w_l D_{KL}(\mathbf{R}_l, \hat{\mathbf{R}}_l) + \text{const}. \end{aligned} \quad (5)$$

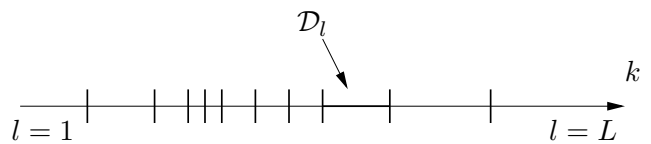


Fig. 3. Partition of the spectral domain to sub-intervals. In each interval  $\mathcal{D}_l$ , the source spectra is constant. The maximum likelihood criteria is a matching between the spectral covariances.

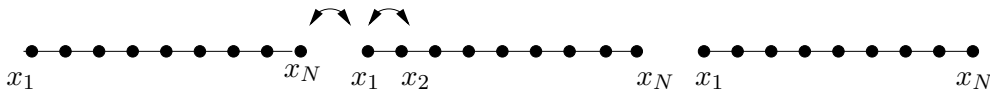


Fig. 2. The circular approximation consists in circularizing the vector  $x_{1..N}$ . For instance, the 1-lag correlation  $r_1 = E[X_1 X_2]$  is equal to  $E[X_N X_1]$ .

The maximum *a posteriori* criteria can be considered, in turn, as a penalized version of the spectral covariance matching:

$$\log p(\boldsymbol{\theta} | \mathbf{x}_{1..K}) = \underbrace{-\sum_{l=1}^L w_l D_{KL}(\mathbf{R}_l, \hat{\mathbf{R}}_l)}_{\text{Matching of spectral covariances}} + \underbrace{\log p(\boldsymbol{\theta})}_{\text{Regularization}} + \text{const.} \quad (6)$$

This separation method is then based on spectral second order statistics. It exploits the spectral non stationarity to estimate the mixing matrix. It is a generalization of the approach adopted in [20] where the authors consider the noiseless case. However, the specificity of the likelihood-based formulation is the exploitation of the hidden structure of the mixture model in order to use the EM algorithm. Therefore, it is based on a statistical approach and inherits the nice properties of the maximum likelihood estimator as the consistency and the efficiency in the asymptotic regime.

### III. NON STATIONARY CASE

Some of the real signals collected in fMRI imaging are obviously non stationary. The difficulties thus arising when applying the spectral matching algorithm [23] to separate the different **temporal** brain activations are the following:

1. The observations are mixture of two types of sources: stationary sources (task related activations, thermal noise,...) and non stationary sources (artifacts).
2. The moderate time duration of scanning, in some experiments, within the limited spectral information provided by time courses make the blind mixing identification a difficult task. In fact, the success of blind separation relies on a good spectral estimation since it is based on structuring the observation covariance matrices according to the linear mixing model.

In order to alleviate the difficulties mentioned above, we have extended the spectral EM algorithm to deal with time varying source spectra. A suitable partition of the time frequency domain leads to a consistent regularization of the estimated spectra. The regularization is performed by windowing the estimated periodograms. Within this extension, a simple EM algorithm is proposed and called hereafter the "Regularized Spectral EM". It keeps a simple form as in [24] and has a double interpretation as the matching of smoothed spectral covariances in the Kullback-Leibler metric and as the maximum likelihood solution when we adopt a Short Time Fourier Transform (STFT) representation of the signals.

#### A. Time dependent spectra

In the non stationary case, the correlation function depends on two variables and its general form is (in the following, we assume that the signals are zero mean):

$$r_x(t_1, t_2) = E[x_{t_1} x_{t_2}] = \int P_x(\omega_1, \omega_2) e^{j(\omega_1 t_1 - \omega_2 t_2)} d\omega_1 d\omega_2,$$

where  $P_x(\omega_1, \omega_2)$  also represents the correlation of the spectral increments which are no more uncorrelated.

The Wigner-Ville spectrum, defined as follows:

$$W_x(t, \omega) = \int_{-\infty}^{\infty} r_x(t + \tau/2, t - \tau/2) e^{-j\omega\tau} d\tau, \quad (7)$$

is an intermediate representation between  $r_x$  and  $P_x$  which mixes time and frequency and thus reflects the notion of time dependent spectrum. In addition, the Wigner-Ville spectrum has many advantages as covariance to time and frequency shifts, real values range and marginal properties [27]:

$$\begin{aligned} \int_{-\infty}^{\infty} W_x(t, \omega) dt &= P_x(\omega, \omega), \\ \int_{-\infty}^{\infty} W_x(t, \omega) d\omega &= r_x(t, t) = \text{var}\{x(t)\}, \end{aligned} \quad (8)$$

Based on the only shift covariance property, a more general class of time dependent spectrum generalizing the Cohen's Class for random signals is:

$$C_x(t, \omega, \phi) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(t-u, \omega-v) W_x(u, v) du \frac{dv}{2\pi} \quad (9)$$

the convolution of the Wigner-Ville spectrum by the 2-D kernel  $\phi$ .

#### B. Maximum Likelihood Separation

The Short Time Fourier Transform (STFT) of a signal  $\{x(t)\}$  is a windowed Fourier transform defined as:

$$S_x(t, \omega) = \int x(\tau) h(\tau - t) e^{-j\omega\tau} d\tau$$

where  $h$  is the moving window capturing the signal non stationarity. It is shown that the squared modulus of  $S_x$  (called the spectrogram) belongs to the Cohen's Class with the kernel  $\phi$  equal to  $W_h$ , the Wigner-Ville distribution of the window  $h$  [27]. Thus, the spectrogram enjoys the positivity property but does not conserve the marginal properties (8) of the Wigner-Ville distribution.

Exploiting the linearity of the STFT transform, the noisy linear mixture model (1) conserves its algebraic form under this transformation:

$$\mathbf{x}(t, \omega) = \mathbf{A}\mathbf{s}(t, \omega) + \mathbf{n}(t, \omega), t = 1..T, \omega = 1..F,$$

where, for the sake of clarity,  $\mathbf{x}$ ,  $\mathbf{s}$  and  $\mathbf{n}$  denote also the STFT transforms of the observations, the sources and the noise respectively. Assuming that the noise is stationary white (with unknown covariance  $\mathbf{R}_\epsilon$ ) and that the sources are decorrelated in the time frequency domain<sup>2</sup> (with unknown diagonal covariances  $\{\mathbf{P}(t, \omega) = \mathbb{E}[\mathbf{s}(t, \omega)\mathbf{s}(t, \omega)^*]\}_{t=1..T}^{\omega=1..F}$ ), the likelihood is as follows:

$$\begin{aligned} p(\mathbf{X} | \boldsymbol{\theta}) &= \int p(\mathbf{X} | \mathbf{S}, \mathbf{A}, \mathbf{R}_\epsilon) p(\mathbf{S} | \{\mathbf{P}(t, \omega)\}) d\mathbf{S} \\ &= \prod_{t, \omega} \int p(\mathbf{x}(t, \omega) | \mathbf{s}(t, \omega), \mathbf{A}, \mathbf{R}_\epsilon) p(\mathbf{s}(t, \omega) | \mathbf{P}(t, \omega)) d\mathbf{s}(t, \omega) \\ &= \prod_{t, \omega} |2\pi \mathbf{R}_{t, \omega}|^{-1} \exp[-\text{Tr}(\mathbf{R}_{t, \omega}^{-1} \mathbf{x}(t, \omega) \mathbf{x}(t, \omega)^*)], \end{aligned} \quad (10)$$

where  $\mathbf{R}_{t, \omega} = \mathbf{A}\mathbf{P}_{t, \omega}\mathbf{A}^* + \mathbf{R}_\epsilon$  and  $\boldsymbol{\theta}$  is the whole parameter to be estimated ( $\mathbf{A}, \mathbf{R}_\epsilon, \{\mathbf{P}(t, \omega)\}$ ).

It is worth noting that the achievement of the separation solution is strongly linked to the diversity of the sources spectrograms (the diagonal time frequency distributions of the matrices  $\mathbf{P}(t, \omega)$  are different). This is the fundamental reason to perform the separation in the time frequency domain when the only spectral and temporal statistics are not able to provide such diversity.

The likelihood (10) can still be interpreted as the matching between STFT covariances matrices  $\mathbf{R}_{t, \omega} = \mathbf{A}\mathbf{P}_{t, \omega}\mathbf{A}^* + \mathbf{R}_\epsilon$  and empirical covariances  $\hat{\mathbf{R}}_{t, \omega} = \mathbf{x}(t, \omega)\mathbf{x}(t, \omega)^*$ , in the Kullback-Leibler metric:

$$\log p(\mathbf{X} | \boldsymbol{\theta}) = - \sum_{t, \omega} D_{KL}(\mathbf{R}_{t, \omega}, \hat{\mathbf{R}}_{t, \omega}) + \text{const}. \quad (11)$$

### C. Time frequency EM algorithm

In spite of its explicit analytic form, the likelihood function (11) is difficult to optimize. However, we can make use of the hidden variable structure of the problem (the sources are the hidden variables) to implement the EM algorithm [25]. The EM algorithm has an iterative scheme. At the iteration  $m$ , it consists of two steps: (E) *Expectation* and (M) *Maximization*.

The first step of the EM algorithm is the computation of the functional  $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m-1)})$  as the *a posteriori* expectation of the complete log-likelihood  $\log p(\mathbf{X}, \mathbf{S} | \boldsymbol{\theta})$ :

$$\begin{aligned} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m-1)}) &= \mathbb{E}[\log p(\mathbf{X}, \mathbf{S} | \boldsymbol{\theta}) | \mathbf{X}, \boldsymbol{\theta}^{(m-1)}] \\ &= \mathbb{E}[\sum_{t, \omega} -\log |\mathbf{R}_\epsilon| - \text{Tr}(\mathbf{R}_\epsilon^{-1} [\mathbf{x}_{t, \omega} - \mathbf{A}\mathbf{s}_{t, \omega}][\mathbf{x}_{t, \omega} - \mathbf{A}\mathbf{s}_{t, \omega}]^*) \\ &\quad + \sum_{t, \omega} -\log |\mathbf{P}_s(t, \omega)| - \text{Tr}(\mathbf{P}_s^{-1}(t, \omega) [\mathbf{s}_{t, \omega}][\mathbf{s}_{t, \omega}]^*)] \end{aligned}$$

where the expectation is computed according to the *a posteriori* distribution of the sources ( $p(\mathbf{S} | \mathbf{X}, \boldsymbol{\theta}^{(m-1)})$ ).

<sup>2</sup>The decorrelation assumption of the time frequency source points is only statistically valid for underspread signals, i.e. the ambiguity function is concentrated in a small neighborhood of the origin [28]. However, our main objective is the estimation of the unknown parameters and not the filtering of sources.

Defining the following statistics which will be computed later:

$$\begin{cases} \mathbf{R}_{xs}(t, \omega) &= \mathbf{x}_{t, \omega} \mathbb{E}[\mathbf{s}_{t, \omega} | \mathbf{x}_{t, \omega}, \boldsymbol{\theta}^{(m-1)}]^* \\ \mathbf{R}_{ss}(t, \omega) &= \mathbb{E}[\mathbf{s}_{t, \omega} \mathbf{s}_{t, \omega}^* | \mathbf{x}_{t, \omega}, \boldsymbol{\theta}^{(m-1)}] \end{cases} \quad (12)$$

the functional  $\mathcal{Q}$  can be rewritten in the following form:

$$\begin{aligned} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m-1)}) &= \sum_{t, \omega} -\log |\mathbf{R}_\epsilon| - \text{Tr}(\mathbf{R}_\epsilon^{-1} [\hat{\mathbf{R}}_{t, \omega} \\ &\quad + \mathbf{A}\mathbf{R}_{ss}(t, \omega)\mathbf{A}^* - \mathbf{A}\mathbf{R}_{sx}(t, \omega) - \mathbf{R}_{sx}^*(t, \omega)\mathbf{A}^*]) \\ &\quad + \sum_{t, \omega} -\log |\mathbf{P}_s(t, \omega)| - \text{Tr}(\mathbf{P}_s^{-1}(t, \omega) \mathbf{R}_{ss}(t, \omega)) \end{aligned} \quad (13)$$

The second step is the update of the parameter  $\boldsymbol{\theta}$  by maximizing the functional  $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m-1)})$ :

$$\boldsymbol{\theta}^{(m)} = \arg \max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m-1)})$$

This can be achieved by differentiating the functional  $\mathcal{Q}$  (13) with respect to the parameter  $\boldsymbol{\theta}$  and then equating to zero. The partial derivatives have the following expression:

$$\begin{aligned} \frac{\partial \mathcal{Q}}{\partial \mathbf{A}} &= -2 \mathbf{R}_\epsilon^{-1} \sum_{t, \omega} (\mathbf{A}\mathbf{R}_{ss}(t, \omega) - \mathbf{R}_{xs}(t, \omega)) \\ \frac{\partial \mathcal{Q}}{\partial \mathbf{R}_\epsilon} &= - \sum_{t, \omega} [\mathbf{R}_\epsilon^{-1} - \mathbf{R}_\epsilon^{-1} (\hat{\mathbf{R}}_{t, \omega} + \mathbf{A}\mathbf{R}_{ss}(t, \omega)\mathbf{A}^* \\ &\quad - \mathbf{A}\mathbf{R}_{sx}(t, \omega) - \mathbf{R}_{sx}^*(t, \omega)\mathbf{A}^*) \mathbf{R}_\epsilon^{-1}] \\ \frac{\partial \mathcal{Q}}{\partial \mathbf{P}_s} &= -(\mathbf{P}_s^{-1} - \mathbf{P}_s^{-1} \mathbf{R}_{ss}(t, \omega) \mathbf{P}_s^{-1}) \end{aligned}$$

leading to the following simple updating equations:

$$\begin{cases} \mathbf{A}^{(m)} &= \mathbf{R}_{xs} \mathbf{R}_{ss}^{-1} \\ \mathbf{R}_\epsilon^{(m)} &= \mathbf{R}_{xx} - \mathbf{R}_{xs} \mathbf{R}_{ss}^{-1} \mathbf{R}_{sx} \\ \text{diag}(\mathbf{P}_s(t, \omega)) &= \text{diag}(\mathbf{R}_{ss}(t, \omega)) \end{cases} \quad (14)$$

where the matrices  $\mathbf{R}_{xx}$ ,  $\mathbf{R}_{xs}$  and  $\mathbf{R}_{ss}$  are the average of the statistic matrices  $\hat{\mathbf{R}}_{t, \omega}$ ,  $\mathbf{R}_{xs}(t, \omega)$  and  $\mathbf{R}_{ss}(t, \omega)$  defined in (12), over the time frequency domain. The matrix  $\mathbf{P}_s(t, \omega)$  is diagonal.

The computation of the statistic matrices (12) is essentially based on the computation of the *a posteriori* first and second moments of the source vector  $\mathbf{s}_{t, \omega}$ . Thanks to the *a priori* Gaussianity of sources and noise, the *a posteriori* distribution of the sources is also Gaussian with the following moments:

$$\begin{cases} \mathbb{E}[\mathbf{s}_{t, \omega}] &= \mathbf{W}_{t, \omega} \mathbf{x}_{t, \omega} \\ \mathbb{E}[\mathbf{s}_{t, \omega} \mathbf{s}_{t, \omega}^*] &= \mathbf{V}_{t, \omega} + \mathbb{E}[\mathbf{s}_{t, \omega}] \mathbb{E}[\mathbf{s}_{t, \omega}]^* \end{cases}$$

where the matrices  $\mathbf{W}_{t, \omega}$  (Wiener matrix) and  $\mathbf{V}_{t, \omega}$  (*a posteriori* covariance) have the following expressions [23]:

$$\begin{cases} \mathbf{W}_{t, \omega} &= [\mathbf{A}^* \mathbf{R}_\epsilon^{-1} \mathbf{A} + \mathbf{P}_s^{-1}(t, \omega)]^{-1} \mathbf{A}^* \mathbf{R}_\epsilon^{-1} \\ \mathbf{V}_{t, \omega} &= [\mathbf{A}^* \mathbf{R}_\epsilon^{-1} \mathbf{A} + \mathbf{P}_s^{-1}(t, \omega)]^{-1} \end{cases}$$

We note that the equations are very similar to a time frequency Wiener filtering. Consequently, the EM algorithm involves an implicit denoising procedure when computing the first *a posteriori* moment of the sources. In other words, we have an optimal source reconstruction at each step of the algorithm.

#### D. Regularized Spectral Matching

The estimation of the parameter  $\theta$  involves the estimation of the whole spectrograms ( $\{\mathbf{P}_s(t, \omega)\}_{t=1..T}^{\omega=1..F}$ ) which are smoothed versions of the Wigner-Ville spectra. In order to accelerate the EM algorithm, we can generalize the procedure presented in the previous Section II to partition the time frequency domain into  $L$  sub-domains  $\{\mathcal{D}_l\}_{l=1}^L$  and then estimate the averaged spectrograms inside these domains. This is algorithmically equivalent to assume that the spectrograms are constant in the sub-domains in the partitioned time frequency 2-D field. The partition is based on some *a priori* information about the signals under hand. An interesting particular case is partitioning the time frequency plane into horizontal bands. Figure 4 illustrates a general partitioning and the segmentation of the time frequency domain into horizontal bands. In the following, we show how the expressions of the updating equations of the EM algorithms are very suitable for such approximation (for a general partition).

We assume that  $\mathbf{P}_s(t, \omega) = \mathbf{P}_l$  for all  $(t, \omega) \in \mathcal{D}_l$  and the likelihood function can be rewritten as follows:

$$p(\mathbf{X} | \theta) = \prod_{l=1}^L |2\pi \mathbf{R}_l|^{-w_l} \exp -\text{Tr} \left( \mathbf{R}_l^{-1} \sum_{(t, \omega) \in \mathcal{D}_l} \mathbf{x}_{t, \omega} \mathbf{x}_{t, \omega}^* \right) \quad (15)$$

where  $w_l = |\mathcal{D}_l|$  and  $\mathbf{R}_l = \mathbf{A} \mathbf{P}_l \mathbf{A}^* + \mathbf{R}_\epsilon$  which is constant in the sub-domain  $\mathcal{D}_l$ .

As the spectral coefficients are constant in each sub-interval  $\mathcal{D}_l$ , the statistics are easily computed. In fact, the matrices  $\mathbf{V}_{t, \omega} = \mathbf{V}_l$  and  $\mathbf{W}_{t, \omega} = \mathbf{W}_l$  are constant over each sub-domain  $\mathcal{D}_l$  and the statistics are:

$$\begin{cases} \hat{\mathbf{R}}(l) &= \frac{1}{w_l} \sum_{(t, \omega) \in \mathcal{D}_l} \mathbf{x}_{t, \omega} \mathbf{x}_{t, \omega}^* \longrightarrow \text{computed off line} \\ \mathbf{R}_{xs}(l) &= \hat{\mathbf{R}}(l) \mathbf{W}_l^* \\ \mathbf{R}_{ss}(l) &= \mathbf{W}_l \hat{\mathbf{R}}(l) \mathbf{W}_l^* + \mathbf{V}_l \end{cases} \quad (16)$$

Then, the statistic matrices  $\mathbf{R}_{xx}$ ,  $\mathbf{R}_{xs}$  and  $\mathbf{R}_{ss}$  are the weighted sums of  $\hat{\mathbf{R}}(l)$ ,  $\mathbf{R}_{xs}(l)$  and  $\mathbf{R}_{ss}(l)$  (with weights  $\{w_l\}_{l=1..L}$ ) and both the mixing matrix  $\mathbf{A}$  and the noise covariance  $\mathbf{R}_\epsilon$  are still updated according to the equation (14). The diagonal sources spectrograms  $\mathbf{P}_l$  are updated according to the following equation:

$$\text{diag}(\mathbf{P}_l) = \text{diag}(\mathbf{R}_{ss}(l))$$

*Remark 1:* An interesting property of the EM algorithm is the fact that the computation of the statistic matrices  $\mathbf{R}_{xs}(l)$  and  $\mathbf{R}_{ss}(l)$  relies on an off-line computation of the

observations spectrograms  $\hat{\mathbf{R}}(l)$ . The computation of the *a posteriori* expectation of sources is no more necessary leading to a fast implementation of the EM algorithm.

Partitioning the time frequency domain into horizontal bands, the algorithm implementing the matching of the STFT spectra has the same structure as in [23] but with covariances computed from the projection of the spectrograms onto the spectral axis. In fact, the projection of the STFT spectrum yields the windowed power spectrum. In Figure 5, we have plotted an fMRI time course in time, its spectrum, its STFT transform and the windowed periodogram (projection of the spectrogram) illustrating the effect of smoothing the spectrum of an fMRI time course. Maximizing the likelihood is then equivalent to matching the windowed periodograms according to equation (11). Thus, the method will essentially consist in maximizing the likelihood of the parameters based on the Gaussian modeling of the sources.

As the computation of the observations spectra is performed off-line, the structure of the algorithm is independent of the partition choice. In fact, the algorithm is only based on matching the computed matrices to structured matrices according to the mixture model. Hereafter the pseudo code of the Regularized Spectral EM algorithm:

**Regularized Spectral EM**

```

1: Initializing:
2: Fixing a partition  $\{\mathcal{D}_l\}_{l=1}^L$ 
3: Off line computation of the smoothed
   covariances  $\hat{\mathbf{R}}(l)$ 
4: Initial values for  $\mathbf{A}$ ,  $\mathbf{R}_\epsilon$  and  $\mathbf{P}_l$ 
5: repeat until convergence,
6: //----- E-step -----//
7: computation of statistics for  $l=1$  to  $L$ ,
8:  $\mathbf{V}_l = (\mathbf{A} \mathbf{R}_\epsilon^{-1} \mathbf{A}^* + \mathbf{P}_l^{-1})^{-1}$ ,  $\mathbf{W}_l = \mathbf{V}_l \mathbf{A}^* \mathbf{R}_\epsilon^{-1}$ 
9:  $\mathbf{R}_{xs}(l) = \hat{\mathbf{R}}(l) \mathbf{W}_l^*$ 
10:  $\mathbf{R}_{ss}(l) = \mathbf{W}_l \hat{\mathbf{R}}(l) \mathbf{W}_l^* + \mathbf{V}_l$ 
11: end of loop on  $l$ ,
12:  $\mathbf{R}_{xs} = \frac{1}{L} \sum w_l \mathbf{R}_{xs}(l)$ 
13:  $\mathbf{R}_{ss} = \frac{1}{L} \sum w_l \mathbf{R}_{ss}(l)$ 
14: //----- M-step -----//
15:  $\mathbf{A} = \mathbf{R}_{xs} \mathbf{R}_{ss}^{-1}$ 
16:  $\mathbf{R}_\epsilon = \text{diag}(\mathbf{R}_{xx} - \mathbf{R}_{xs} \mathbf{R}_{ss}^{-1} \mathbf{R}_{xs}^*)$ 
17:  $\mathbf{P}_l = \text{diag}(\mathbf{R}_{ss}(l))$ , for  $l=1$  to  $L$ 
18: Renormalize  $\mathbf{A}$  and  $\mathbf{P}_l$ 
19: end of repeat

```

(17)

#### IV. ILLUSTRATION ON FMRI DATA

The Regularized Spectral EM algorithm was applied to separate the time courses of fMRI data acquired at the FM Kirby Center for Functional Brain Imaging. The experiment consisted of presenting two periodic visual stimulus, shifted by 20 s from one another, to the subject. The stimuli consisted of an 8-Hz reversing checkerboard pattern presented for 15 s in the right visual hemifield, followed by 5 s of an asterisk fixation, followed by 15 s of checkerboard presented to the left visual hemifield, followed by 20 s of an asterisk fixation. The 55 s set of events was repeated

four times for a total of 220 s. Scans were acquired on a Philips NT 1.5-Tesla scanner. A sagittal localizer scan was performed first, followed by a  $T_1$ -weighted anatomic scan [repeat time ( $TR$ ) = 500 ms, echo time ( $TE$ ) = 30 ms, field of view = 24 cm, matrix =  $256 \times 256$ , slice thickness = 5 mm, gap = 0.5 mm] consisting of 18 slices through the entire brain including most of the cerebellum. Next, we acquired functional scans over the same 18 slices consisting of a single-shot, echo-planar scan ( $TR$ =1 s,  $TE$ = 39 ms, field of view = 24 cm, matrix=  $64 \times 64$ , slice thickness = 5 mm, gap = 0.5 mm, flip angle = 90 degrees) obtained consistently over a 3-min, 40-s period for a total of 220 scans.

Our method is tested on a single-subject data. It is applied on two different slices where we expect, in each, two different task related components corresponding to the alternating activation of the right and left visual cortex as a response to an alternating visual stimulus presented to the subject. The temporal fMRI separation relies on the following mixing model:

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{N}$$

where  $\mathbf{X}$  is the ( $M \times T$ ) matrix of observations, the column  $\mathbf{X}(:, t)$  contains the scanned image acquired at the time  $t$  and  $M$  is the number of voxels in one brain slice. The ( $N \times T$ )-matrix  $\mathbf{S}$  contains the  $N$  time courses rows. The matrix  $\mathbf{N}$  models the noise corrupting the observations. The advantage of taking into account the noise in the model is to allow the separating algorithm to only extract the relevant components. This is possible when the spectral profile of the noise is flat comparing to the more concentrated component spectra. In fact, the time courses of task related activations present the same frequency content as the stimulus.

The results of the proposed algorithm applied to the first data set (slice 10) are shown in Figure 6, where we have plotted three recovered image components (the three columns of the estimated mixing matrix  $\hat{\mathbf{A}}$ ) within their corresponding estimated time courses. We note the ability of the algorithm to extract the components which have a time course correlated with the stimulus. The first and third components correspond to the alternative activations of the right and left visual cortex as expected from the conditions of the stimulus presented to the subject. However, the algorithm extracts also another component (the second one) which has a spectral density similar to the first two components. This shows that fixing the number of components by intuitive expectation (2 components) based on the experiment paradigm leads to wrong results. In fact, fixing the number of components to 2, the Regularized EM algorithm yields a component consisting of a mixture of the right visual cortex activation and a transient signal. The right visual cortex activation is not thus well extracted. The results shown in Figure 6 were obtained by varying the number of components and then studying *a posteriori* the results after convergence of the EM algorithm. An automatic selection of the number of components is thus needed for a complete blind analysis of the fMRI data. We discuss

this point later in Section VI. Figure 7 illustrates the time courses of the right and left visual cortex regions. We note the periodicity of the time courses and their relative inter-delay (around 20 s) corresponding to the inter-delay of the stimulus (the checkerboard pattern was presented alternatively to the right and left visual hemifields). Figure 8 shows the spectrograms and the regularized spectra of the estimated three time courses. For comparison purposes, we reported the separation results of a temporal ICA InfoMax algorithm on the same data set, in Figure 9. We have fixed the number of components to 3. The ICA algorithm fails to extract the third component from the two alternative task related components identified with the EM algorithm, mixing them with a higher frequency component. In order to take into account the noise component in the InfoMax method, we have also fixed the number of components to 4. Figure 10 illustrates the extracted components. The first and fourth components correspond to the left and right visual cortex activations. However, the remaining two components contain a mixture of the transient signal and the right visual cortex activations. The noise is not extracted in an independent component but seems to be spread in components 2, 3 and 4.

In Figure 11, the results are plotted for the slice 14 characterized by a higher noise level than the previous slice. We note the ability of the algorithm to recover the components as well. For this slice, we have processed only 110 timepoints as the acquisition rate was divided by 2 in the same experiment conditions as the data acquisition for the slice 10. Figure 12 illustrates the estimated time courses. We note their correlation with the expected activations of the brain as a response to the presented stimulus. The time delay is about 10 timepoints corresponding to the 20 s of the shift between the alternative periodic right and left visual stimulus. The ability of the algorithm to extract the components is then unaltered by dividing the sample size by two. The InfoMax ICA algorithm gives poor results (Figure 13) essentially because of the presence of noise in the processed data. This is to be expected as, in the ICA solution, we assume a noiseless observation model.

## V. CONCLUSION

We have presented the extension of the Spectral EM algorithm to deal with real data suffering from non stationarity and a lack of enough points for spectral analysis. The separation method is essentially based on the diversity of the source smoothed periodograms. The non stationarity of second order statistics allows the mixing matrix identification without resorting to higher order statistics. The use of second order statistics (in other words, the Gaussian modeling) leads to an efficient and fast implementation of the EM algorithm. In fMRI data analysis, we have exploited this diversity between the time course spectra. The spatial pixel distributions represent the columns of the mixing matrix. Therefore, the Regularized Spectral EM algorithm allows a blind joint estimation of the brain source images within the spectra of their time courses. The task related sources are easily distinguished by the signature of

the stimulus in their time course spectra.

## VI. PERSPECTIVES

We have not yet exploited another powerful feature of the Bayesian approach which is the incorporation of prior information into the inference process. Hereafter, we briefly outline how to incorporate such prior information.

As our method is based on the identification of the spectral structure of the covariance matrices, one can take advantage of some prior information about the task related time course spectrum. In fact, this later reflects the spectral information of the stimulus presented to the subject during the scanning. This can be consistently performed by constructing a prior function  $p(\mathbf{P}_k)$  and then optimize the penalized criteria:

$$\begin{aligned} \log p(\mathbf{A}, \mathbf{P}_k, \mathbf{R}_\epsilon | \mathbf{x}_{1..N}) = & \\ & \underbrace{-\sum_{l=1}^L w_l D_{KL}(\mathbf{R}_l, \hat{\mathbf{R}}_l)}_{\text{Matching of spectral covariances}} + \underbrace{\log p(\mathbf{P}_k)}_{\text{Regularization}} + \text{const.} \end{aligned} \quad (18)$$

In a previous work [29], we have presented a theoretical rule to automatically construct an uninformative prior based on information geometry [30]. Let  $\mathbf{P}_{ref}$  a reference spectrum (e.g. that of the stimulus), an uninformative prior in the 0-geometry is an Inverse Wishart prior on the spectra  $\mathbf{P}$ :

$$\begin{aligned} p(\mathbf{P} | \mathbf{P}_{ref}) &= \mathcal{IW}(\mathbf{P}; \nu, \mathbf{P}_{ref}) \\ &\propto |\mathbf{P}|^{-\frac{\nu-(n+1)}{2}} \exp\left[-\frac{\nu}{2} \text{Tr}(\mathbf{P}^{-1} \mathbf{P}_{ref})\right] \end{aligned}$$

where  $\nu$  is the degree of freedom and reflects the confidence degree in the reference spectra. An important feature of the Inverse Wishart prior is the fact that the update equations of the EM algorithm keep the same form as in (17). Consequently, the maximum *a posteriori* solution (minimization of the penalized cost function) shares the same implementation efficiency of the maximum likelihood solution. We are testing this modified version and results will be reported in a forthcoming paper.

In the proposed method, we have fixed the number of sources  $n$  according to what we expect from the specificity of the stimulus presented to the subject during the experiment. For an unconstrained analysis, this subjective choice is considered as a limitation. The intuitive expectation from the experiment paradigm could be wrong. For instance, in the experiment presented in Section IV, we have expected two relevant sources matching the alternating left and right stimulus presented to the subject during scanning. However, by varying the number of sources and running the algorithm, we have identified a third task related source (see Figure 6). Thus, an automatic and objective estimation of the number of sources is certainly an important direction of research in Blind Source Separation in general and for fMRI data analysis in particular. Fortunately, the Bayesian framework allows a consistent

incorporation of the number of sources in the problem formulation as a random variable  $n$ . Let  $p(n | \mathcal{I})$  be a prior on the random variable  $n$ .  $\mathcal{I}$  represents all our prior information. A consistent estimation of  $n$  can be performed by maximizing its *a posteriori* marginal probability<sup>3</sup>. The *a posteriori* marginal probability  $p(n | \mathbf{X}, \mathcal{I})$  is obtained by integrating over the parameter  $\boldsymbol{\theta} = \{\mathbf{A}, \mathbf{R}_\epsilon, \{\mathbf{P}_k\}\}$ :

$$\begin{aligned} p(n | \mathbf{X}, \mathcal{I}) &= \int_{\boldsymbol{\theta}} p(n, \boldsymbol{\theta} | \mathbf{X}, \mathcal{I}) d\boldsymbol{\theta} \\ &= \int_{\boldsymbol{\theta}} \int_{\mathbf{S}} p(n, \boldsymbol{\theta}, \mathbf{S} | \mathbf{X}, \mathcal{I}) d\mathbf{S} d\boldsymbol{\theta} \end{aligned} \quad (19)$$

Computation of the integrals in the expression (19) is intractable. However, Bayesian sampling technics such as the Reversible Jump MCMC [31] are well suited to the data augmentation structure of our problem. It consists in sampling the parameter  $\boldsymbol{\theta}$  and the sources  $\mathbf{S}$  in a Gibbs cyclic way with a random additional step of birth or death of a source component. We are still exploring this Bayesian sampling technic.

## ACKNOWLEDGMENT

This work was supported in part by the National Institutes of Health under grant 1 R 01 EB 000840-01 (to VC).

## REFERENCES

- [1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley and Sons, 2001.
- [2] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*, John Wiley, 2003.
- [3] M. McKeown, S. Makeig, G. Brown, T. Jung, S. Kindermann, A. Bell, and T. Sejnowski, "Analysis of fMRI data by blind separation into independent spatial components.", *Hum. Brain Mapp.*, vol. 6, pp. 160–188, 1998.
- [4] V. Calhoun, T. Adali, L. Hansen, J. Larsen, and J. Pekar, "ICA of functional MRI data: An overview", in *Fourth International Symposium on Independent Component Analysis and Blind Source Separation*, Nara, Japan, April 2003, pp. 281–288.
- [5] C. Jutten and J. Herault, "Blind separation of sources. I. an adaptive algorithm based on neuromimetic architecture", *Signal Processing*, vol. 24, no. 1, pp. 1–10, 1991.
- [6] P. Comon, "Independent Component Analysis, a new concept?", *Signal processing, Special issue on Higher-Order Statistics*, Elsevier, vol. 36 (3), pp. 287–314, April 1994.
- [7] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis", *Neural Computation*, vol. 9(7), pp. 1483–1492, 1997.
- [8] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non Gaussian signals", *IEE Proceedings-F*, vol. 140, no. 6, pp. 362–370, December 1993.
- [9] A. J. Bell and T. J. Sejnowski, "An information maximization approach to blind separation and blind deconvolution", *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, 1995.
- [10] E. Moulines, J. Cardoso, and E. Gassiat, "Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models", in *ICASSP-97*, Munich, Germany, April 1997.
- [11] H. Attias, "Independent factor analysis", *Neural Computation*, vol. 11, pp. 803–851, 1999.
- [12] H. Snoussi and A. Mohammad-Djafari, "Fast joint separation and segmentation of mixed images", *Journal of Electronic Imaging*, vol. 13, no. 2, pp. 349–361, April 2004.
- [13] A. Belouchrani, K. Abed Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique based on second order statistics", *IEEE Trans. Signal Processing*, vol. 45, no. 2, pp. 434–444, February 1997.

<sup>3</sup>The estimation of the source number can be considered as a model selection problem



- [14] K. Rahbar and J. Reilly, "Blind source separation of convolved sources by joint approximate diagonalization of cross-spectral density matrices", in *Proc. ICASSP*, 2001.
- [15] A. Belouchrani and M. Amin, "Blind source separation using time-frequency distributions: algorithm and asymptotic performance", in *Proc. ICASSP*, Munchen, 1997, pp. 3469 – 3472.
- [16] E. Weinstein, M. Feder, and A. Oppenheim, "Multi-channel signal separation by decorrelation", *IEEE Trans. Speech, Audio Processing*, vol. 1, no. 4, October 1993.
- [17] K. Matsuoka, M. Ohya, and M. Kawamoto, "A neural net for blind separation of nonstationary sources", *Neural Networks*, vol. 8(3), pp. 411–419, 1995.
- [18] S. Choi and A. Cichocki, "Blind separation of nonstationary sources in noisy mixtures", *Electronics Letters*, vol. 36(9), pp. 848–849, apr 2000.
- [19] A. Souloumiac, "Blind source detection and separation using second order nonstationarity", in *Proc. ICASSP*, 1995, pp. 1912–1915.
- [20] D.-T. Pham and J. Cardoso, "Blind separation of instantaneous mixtures of non stationary sources", *IEEE Trans. Signal Processing*, vol. 49, 9, no. 11, pp. 1837–1848, 2001.
- [21] J. Cardoso, "The three easy routes to independent component analysis; contrasts and geometry", in *Proc. of the ICA 2001 workshop*, December 2001.
- [22] B. R. Hunt, "A matrix theory proof of the discrete convolution theorem", *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 285–288, 1971.
- [23] H. Snoussi, G. Patanchon, J. Macías-Pérez, A. Mohammad-Djafari, and J. Delabrouille, "Bayesian blind component separation for cosmic microwave background observations.", in *Bayesian Inference and Maximum Entropy Methods*, R. L. Fry, Ed. MaxEnt Workshops, August 2001, pp. 125–140, Amer. Inst. Physics.
- [24] J. Cardoso, H. Snoussi, J. Delabrouille, and G. Patanchon, "Blind separation of noisy Gaussian stationary sources. application to cosmic microwave background imaging", in *Eusipco*, Toulouse, France, September 2002.
- [25] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", *J. R. Statist. Soc. B*, vol. 39, pp. 1–38, 1977.
- [26] G. Patanchon, H. Snoussi, J. Cardoso, and J. Delabrouille, "Component separation for cosmic microwave background data: a blind approach based on spectral diversity", in *PSIP*, Grenoble, France, January 2003.
- [27] P. Flandrin and W. Martin, *The Wigner-Ville Spectrum of Nonstationary Random Signals*, Elsevier Science. W. Mecklenbrauker and F. Hlawatsch (Editors), 1997.
- [28] F. Hlawatsch, G. Matz, H. Kirchauer, and W. Kozek, "Time-frequency formulation, design, and implementation of time varying optimal filters for signal estimation", *IEEE Trans. Signal Processing*, vol. 48, no. 5, pp. 1417–1432, May 2000.
- [29] H. Snoussi and A. Mohammad-Djafari, "Information Geometry and Prior Selection", in *Bayesian Inference and Maximum Entropy Methods*, C. Williams, Ed. MaxEnt Workshops, August 2002, pp. 307–327, Amer. Inst. Physics.
- [30] S. Amari and H. Nagaoka, *Methods of Information Geometry*, vol. 191 of Translations of Mathematical Monographs, AMS, OXFORD, University Press, 2000.
- [31] P. J. Green, "Reversible jump MCMC computation and Bayesian model determination", *Biometrika*, vol. 82, pp. 711–732, 1995.

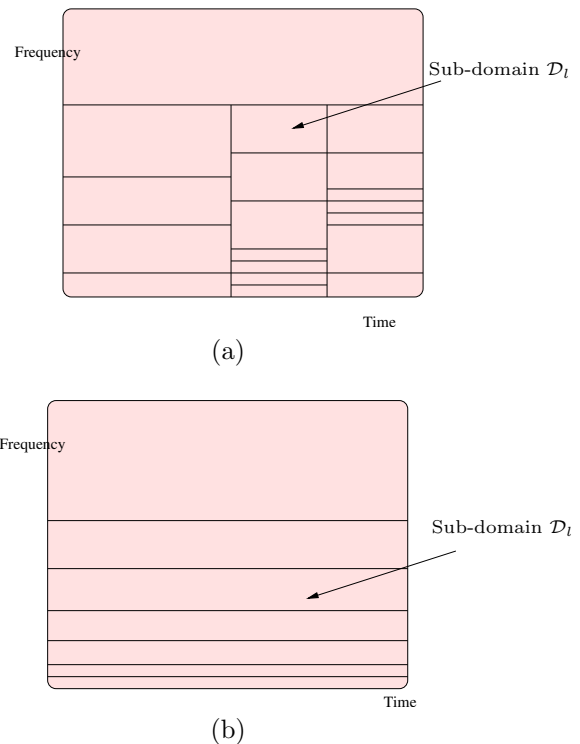


Fig. 4. (a) General partitioning of the time frequency domain, (b) marginal partitioning of the Time Frequency domain into horizontal bands: exploitation of the spectral non stationarity.

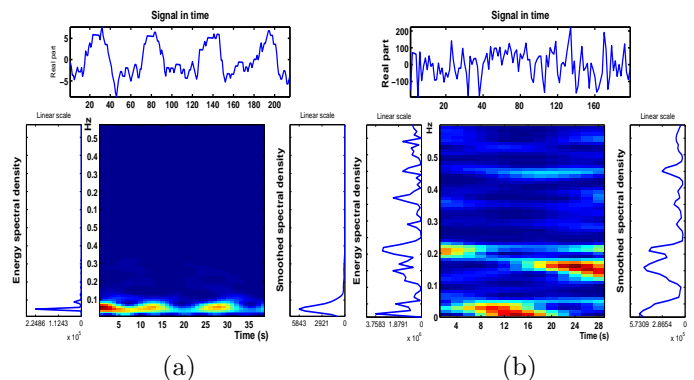


Fig. 5. Time frequency behavior of fMRI time courses (the sample size is 220s, a Hamming window with length 64s and spacing 4s is used, only the half of the frequency range is shown): (a) fMRI time course corresponding to the right visual cortex activation, (b) artifact related fMRI time course. The projection of STFT signal transform has a smoothness effect on the spectrum.

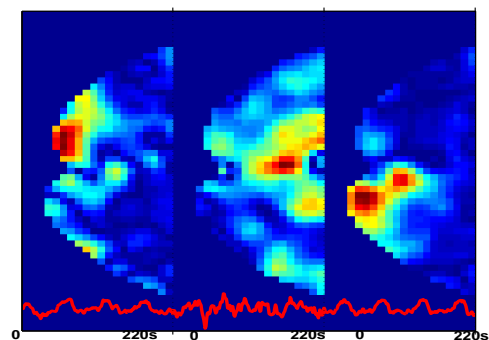


Fig. 6. The recovered components with the Regularized Spectral EM algorithm for slice 10. The first and third components correspond to the left and right visual cortex activations.

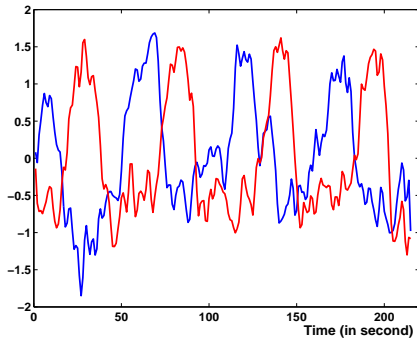


Fig. 7. The recovered time courses with the Regularized Spectral EM algorithm. Their temporal inter-delay is about 20 s corresponding to inter-delay between the alternative temporal stimulus presented to the subject.

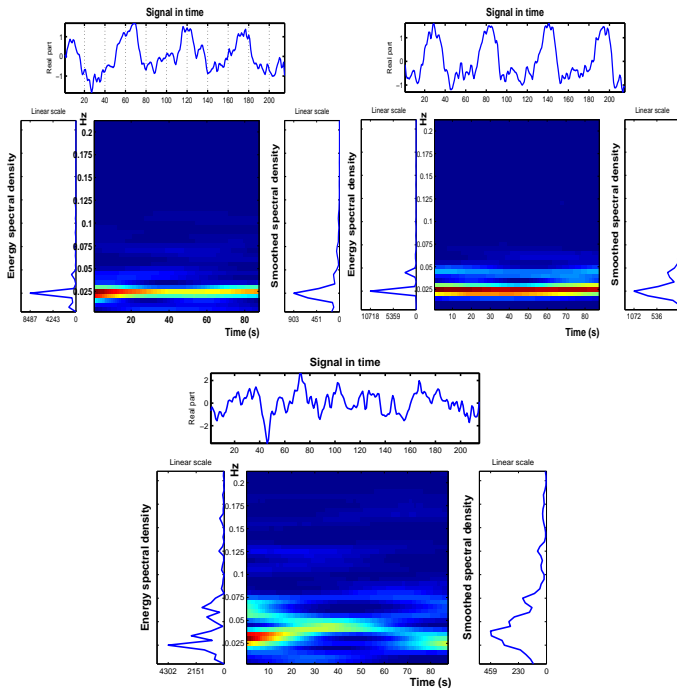


Fig. 8. The spectrograms of the estimated time courses (the sample size is 220s, a Hamming window with length 64s and spacing 1s is used, only the half of the frequency range is shown): First, third and second according to the ordering in Figure 6.

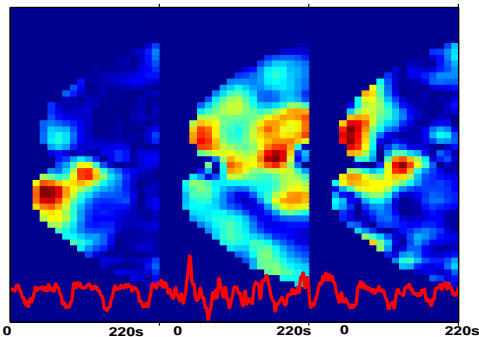


Fig. 9. The recovered components with the ICA InfoMax algorithm for the Slice 10. The alternative task related components are not separated from the transient signal (the image in the middle of the Figure 6).

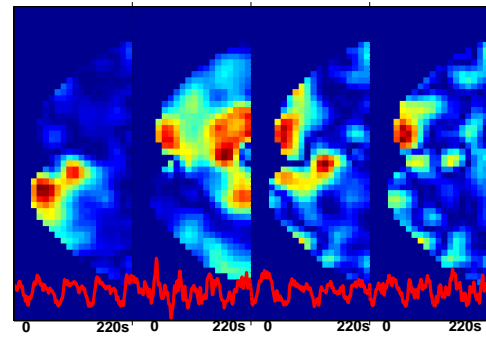


Fig. 10. The recovered components with the ICA InfoMax algorithm for the Slice 10. The number of components is set fixed to 4 in order to take into account the noise.

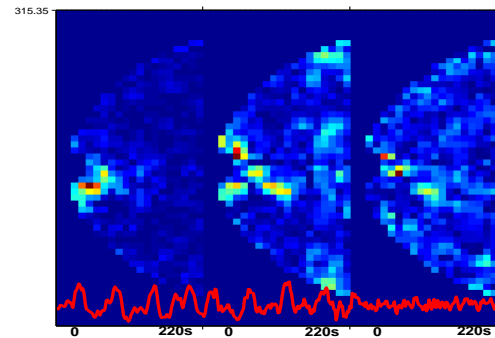


Fig. 11. The recovered components with the Regularized Spectral EM algorithm for the slice 14. The first and second components correspond to the right and left visual cortex.

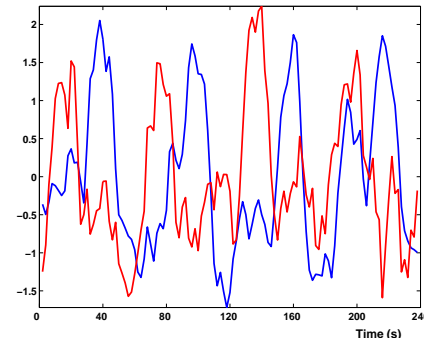


Fig. 12. The recovered time courses with the Regularized Spectral EM algorithm. We note the periodicity and the inter-delay compatible with the periodicity and the inter-delay of the alternative left and right stimulus presented to the subject during scanning.

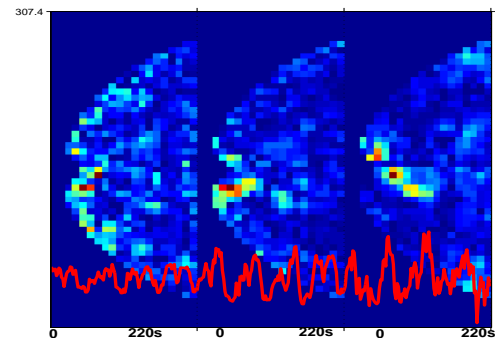


Fig. 13. The recovered components with the ICA InfoMax algorithm for the slice 14 in a high noise environment. The components are not identified.